



Global AI Server Market and Supply Chain Development Forecast for the Next Five Years

Research Operations Center
Department of Semiconductor Research
2024/12



Analysis, opinions and recommendations contained within the report are designed to help our clients make informed decisions promptly based on the comprehensive collection of information and judgment. The information and statistical data herein have been obtained from sources believed to be reliable, but the accuracy and completeness of the information are not guaranteed. TrendForce(集邦科技股份有限公司) makes no guarantee, representation of warranty and accepts no responsibility or liability as to its accuracy or completeness. The information and analysis in the report constitute judgment as at the date of this report and are subject to change without prior notice. TrendForce (集邦科技股份有限公司) is not liable for any incidental, consequential or indirect (including but not limited to) damages for lost profits, business interruption and loss of information arising out of the use of delay of any information provided by TrendForce (集邦科技股份有限公司). All material presented in the report, unless specifically indicated, is under copyright, International Copyright Laws, and other relevant laws to TrendForce (集邦科技股份有限公司). All copyrights, patents, trademarks, business secrets, as well as the property rights, ownership, and other rights of other intellectual properties in the report are owned by TrendForce (集邦科技股份有限公司). Clients have the right to use the information in the report for internal use. However, no part of the report may be modified, copied, republished, reproduced, displayed, transmitted or distributed in any form without prior written consent from TrendForce (集邦科技股份有限公司). Any person breaching the intellectual property rights will have to bear full legal responsibility; TrendForce (集邦科技股份有限公司) may ask the offender to suspend any rights infringement behavior and request for indemnification in accordance with the laws. The contents of this report and any attachments are confidential and legally protected from disclosure.

CONTENT

**2023-2028F Global AI Server Market
Supply and Demand Analysis**

**2023-2028F Forecast for Global AI Server
Market**

**Future Trends and Outlook for Key AI
Server Supply Chain Development**

Key Takeaways

A background image showing a business meeting. In the foreground, a person's hand is pointing at a document with a bar chart and a pie chart. To the left, a pair of glasses and a spiral notebook are visible. In the background, another person is holding a pen and looking at a tablet. The scene is brightly lit, suggesting a professional office environment.

2023-2028F Global AI Server Market Analysis and Forecast

Analysis on Key Factors Impacting 2023-2028F Global AI Server Shipment Development

Estimated Impact Levels

Growth Drivers from the Demand Side

XXXX

★★★

XXXX

★★

XXXX

★★

XXXX

★★

Growth Drivers from the Supply Side

XXXX

★★

XXXX

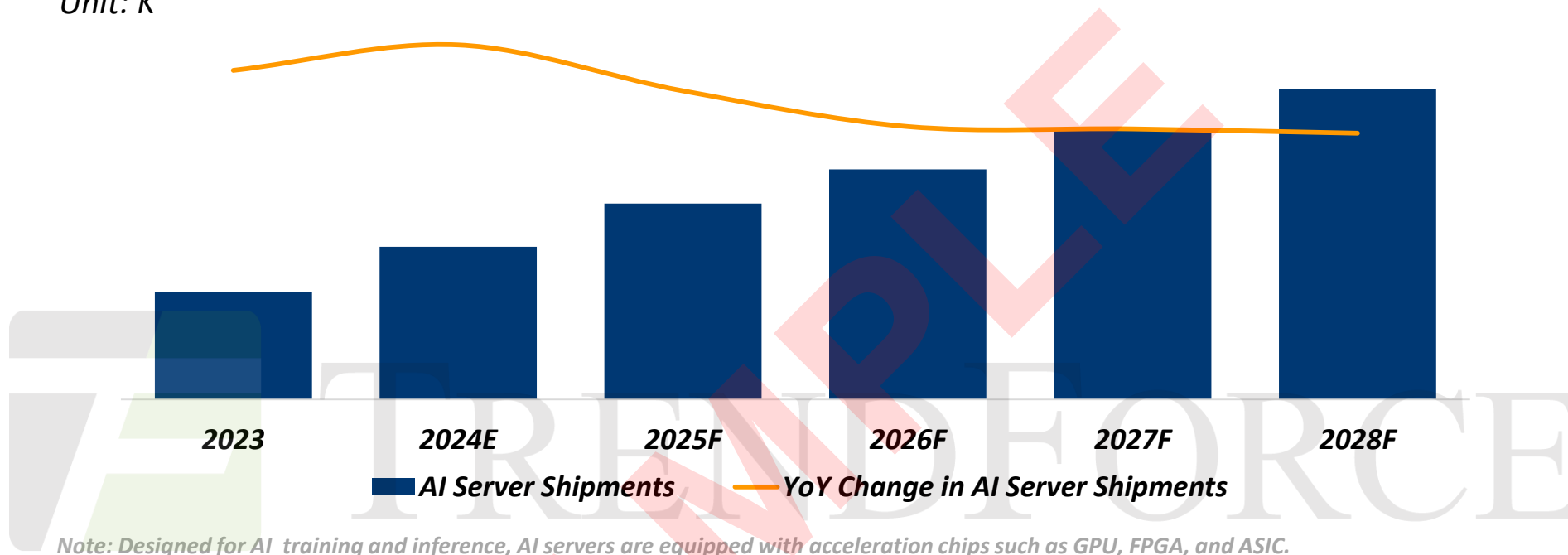
★

Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

Growth Forecast for AI Server Market, 2023-2028F

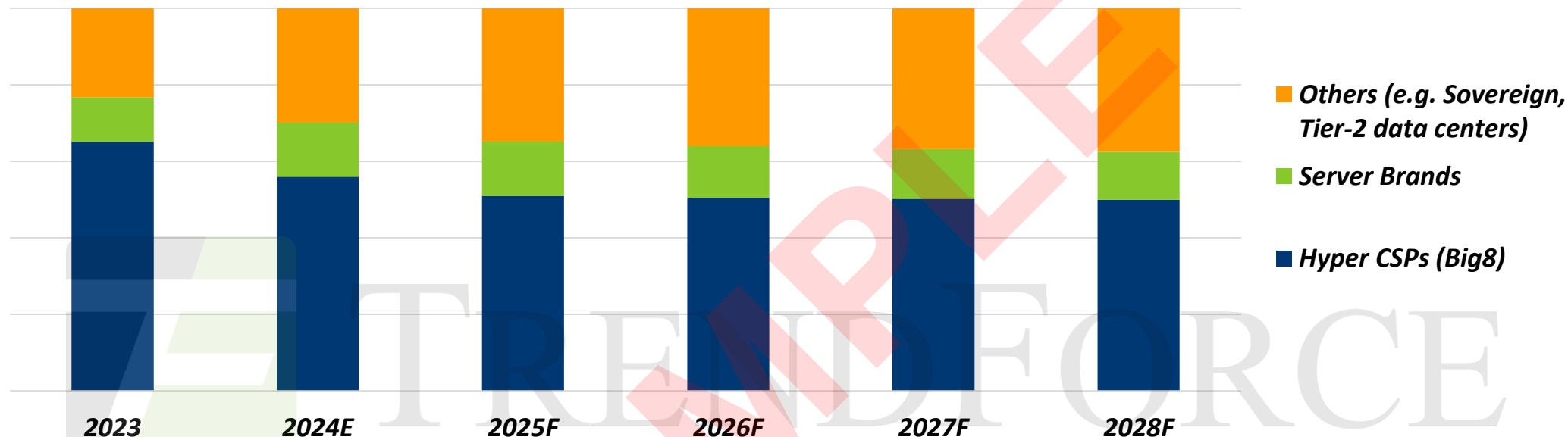
Unit: K



- In response to the rise of GenAI technologies like ChatBot, the market for AI servers will experience a surging growth between 2023 and 2024 at an average YoY growth of roughly XXXX% in shipment...

Hyper CSPs Will Still Represent the Largest Group of Customers for High-end AI Servers by 2028

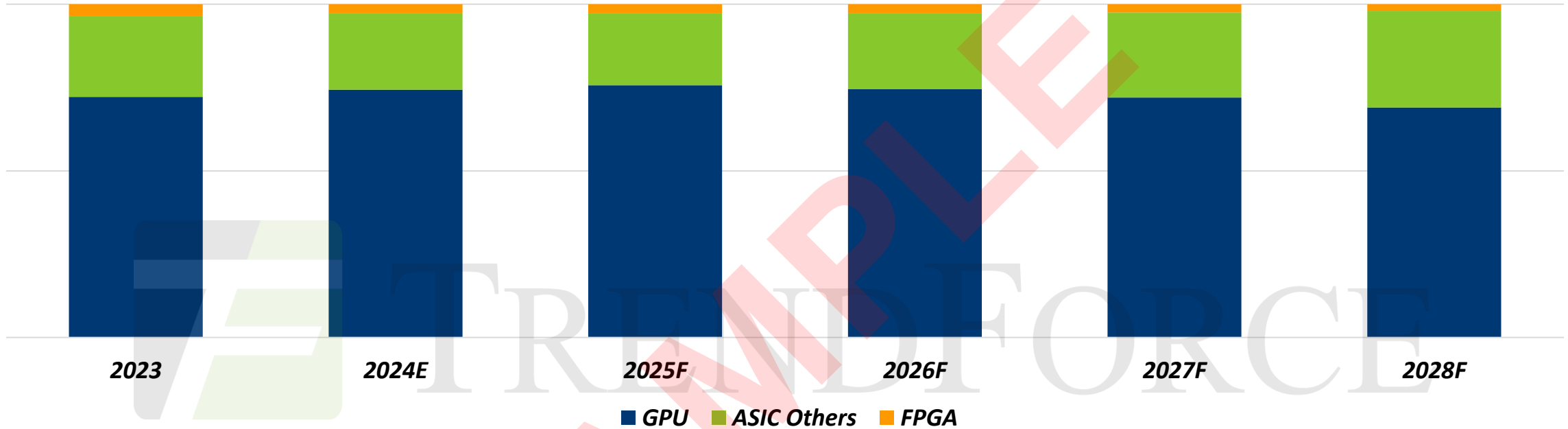
2023-2028F Distribution of High-end AI Server Shipments by Customer Groups



- “Hyper CSPs”, or the world’s eight largest CSPs encompassing Microsoft, AWS, Google, Meta, and BBAT, are the major customers in the AI server market. Hyper CSPs are expected to remain as the largest customer group for purchasing high-end AI servers in 2028, accounting for approximately XXXX% of the market share...

CSPs Would Actively Expand Their Own ASIC AI Server Infrastructure, with The Proportion to XXXX% by 2028

Projected Shipment on AI Servers adopted with Various Chips between 2023 and 2028F

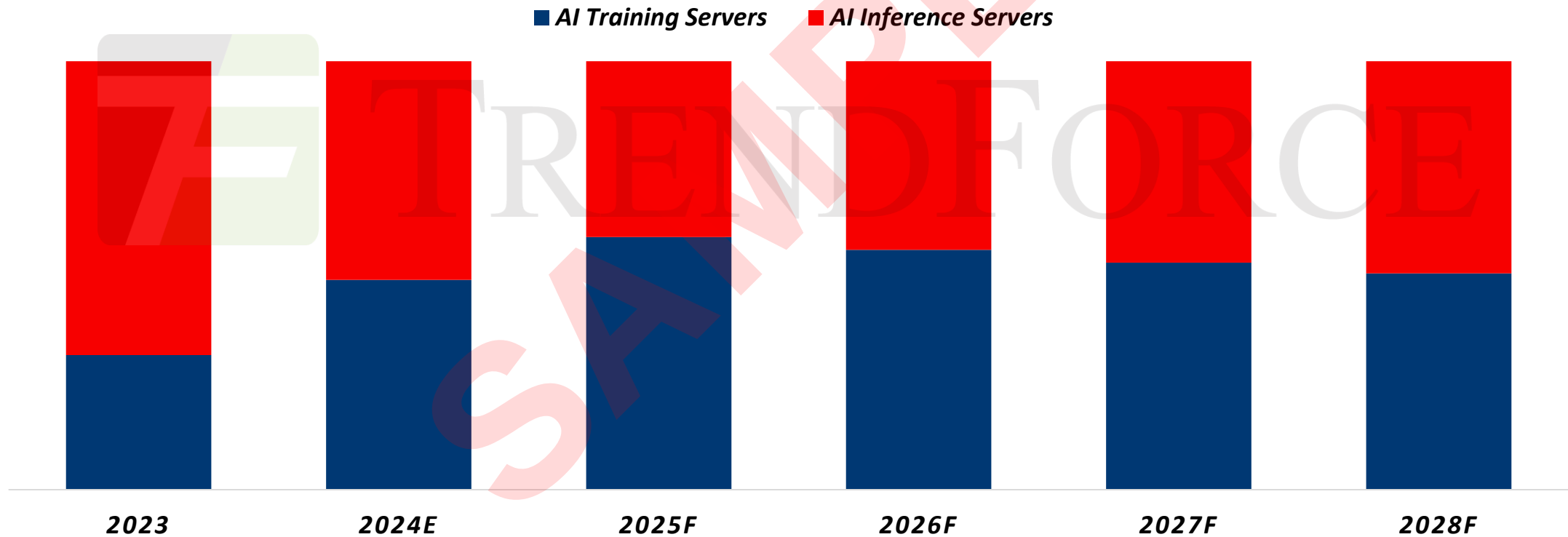


□ NVIDIA remains as the largest supplier as H20 was not included in the sanction imposed by BIS in December 2024...

Proportion of AI Inference Servers Expected to Approach XXXX% by 2028

- The market for AI training servers will experience a rapid growth between 2023 and 2025 due to the rising demand from CSPs and server brands. AI training servers are projected to account around XXXX% of the total AI server shipments in 2025...

Shipment Share of AI Training Servers vs. That of AI Inference Servers, 2023-2028F



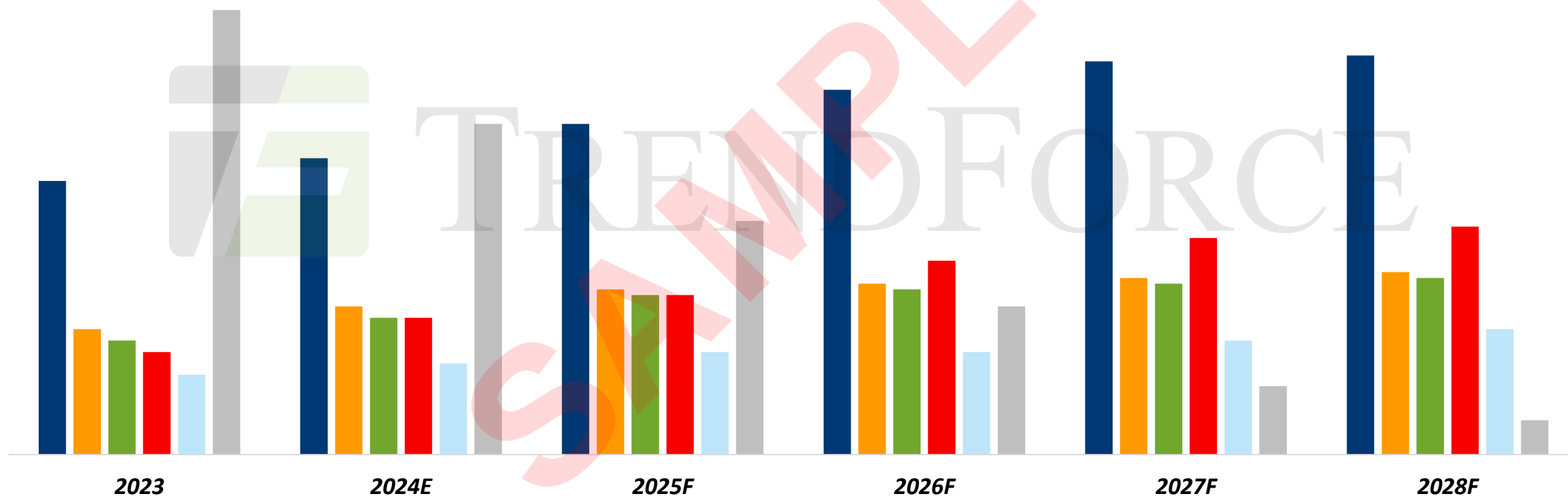
Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

Estimation for Development Dynamics of Major AI Applications

AI applications primarily driven by major CSPs...

Cloud/E-Commerce/ChatBOT Smart Manufacturing Finance/Insurance Autonomous Vehicles/Traffic Smart Healthcare Others



Source: TrendForce, Dec. 2024

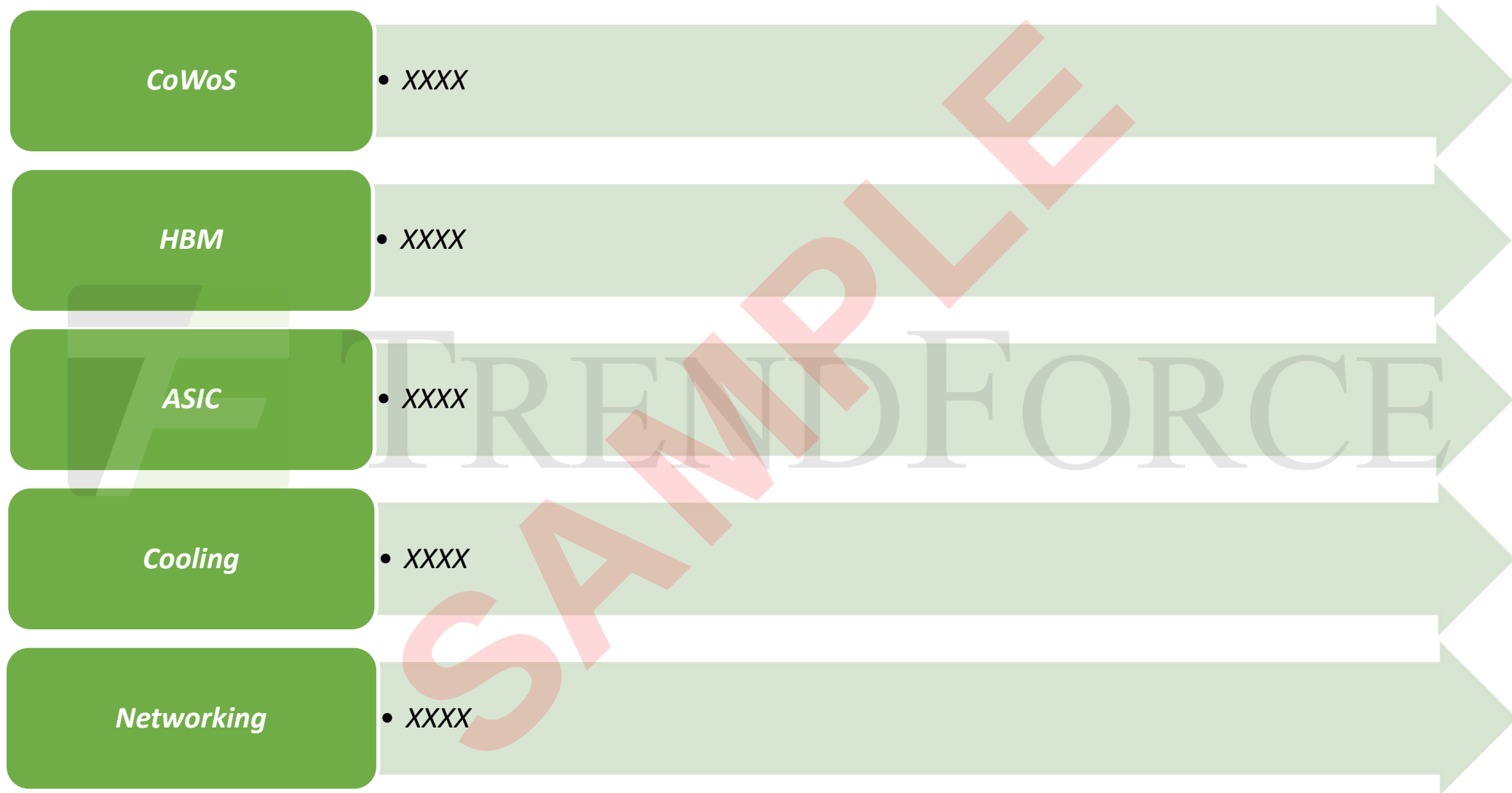
The contents of this report and any attachments are confidential and legally protected from disclosure.

FOR INTERNAL USE ONLY

A background image showing a business meeting. In the foreground, a person's hand is writing on a document with a pen. The document features a bar chart with blue and orange bars, and a pie chart. To the left, a pair of black-rimmed glasses rests on the table. In the background, another person is holding a tablet and a pen, looking at it. There are also some books or folders on the table. The overall scene is brightly lit, suggesting a professional office environment.

Future Trends and Outlook for Key AI Server Supply Chain Development

The Development Trend of AI Server Supply Chain, 2023-2028F



Source: TrendForce, Dec. 2024

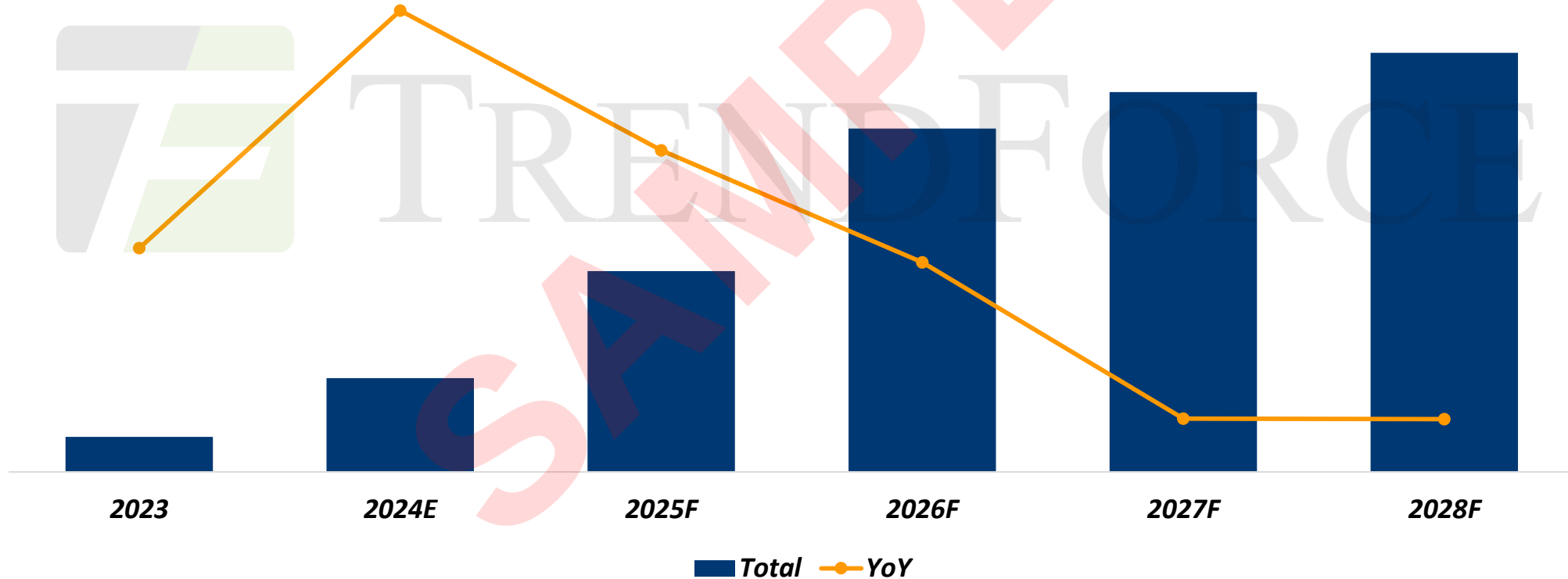
The contents of this report and any attachments are confidential and legally protected from disclosure.

FOR INTERNAL USE ONLY

TSMC's CoWoS Production Capacity Forecast in 2023-2028F

□ TSMC's overall CoWoS capacity in 2025 is expected to reach XXXXK at a growth rate of XXXX%....

Unit: K



Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

FOR INTERNAL USE ONLY

Development of AI Chips and Comparison of HBM Specifications between NVIDIA and AMD

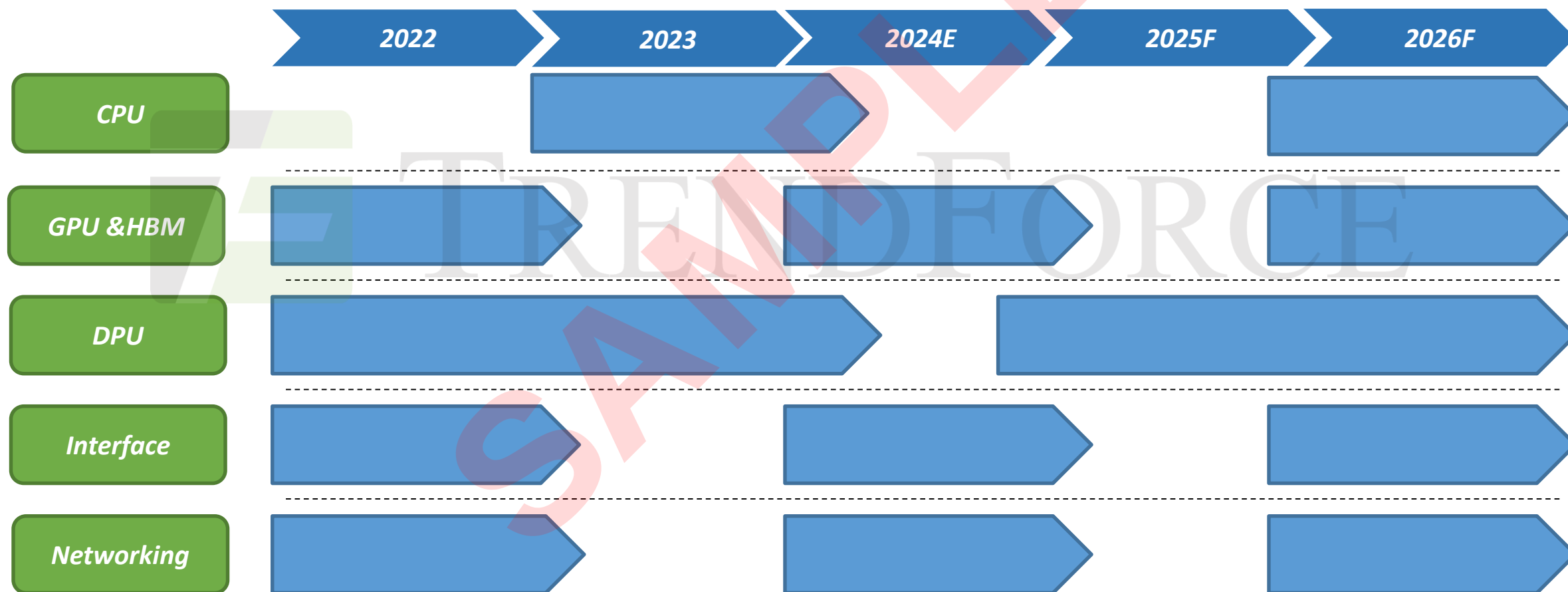
Company	AI Chips	2023				2024E				2025F				2026F			
		1Q23	2Q23	3Q23	4Q23	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26
 NVIDIA	H100																
	GH200 CPU+GPU)																
	H20																
	H200																
	B100/B200																
	GB200 (CPU+GPU)																
	Blackwell Ultra																
	R100																
 AMD	MI200																
	MI300X																
	MI300A (CPU+GPU)																
	MI325X																
	MI350																
	MI400																

Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

NVIDIA GPU, CPU and Related AI solution Portfolio Changes

- NVIDIA typically launches a new platform product, including CPUs and GPUs...

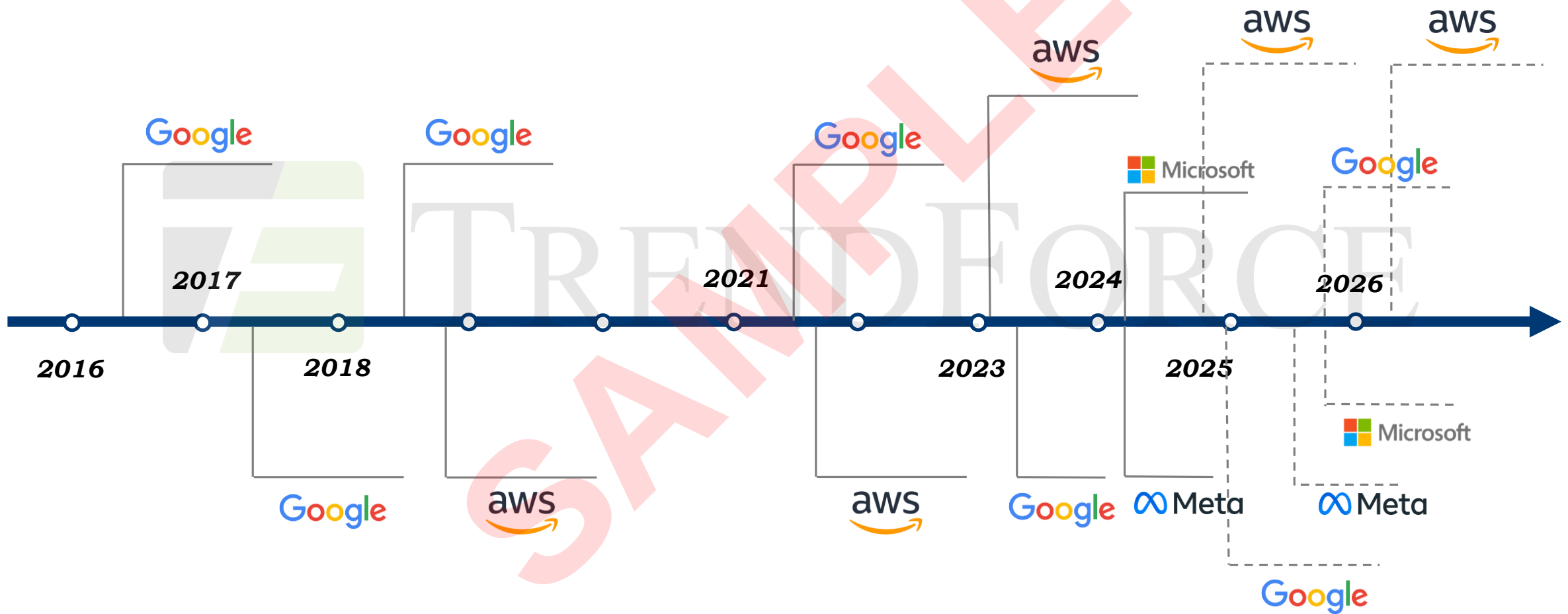


Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

Major US CSPs Keep on Developing In-House AI Chips

- Google is a pioneer in in-house ASIC development among US CSPs, and has been developing its own AI chip TPU since 2016...



Source: TrendForce, Dec. 2024

CSPs Building In-house Chips with More Advanced Technology Node and Higher Bandwidth Memory

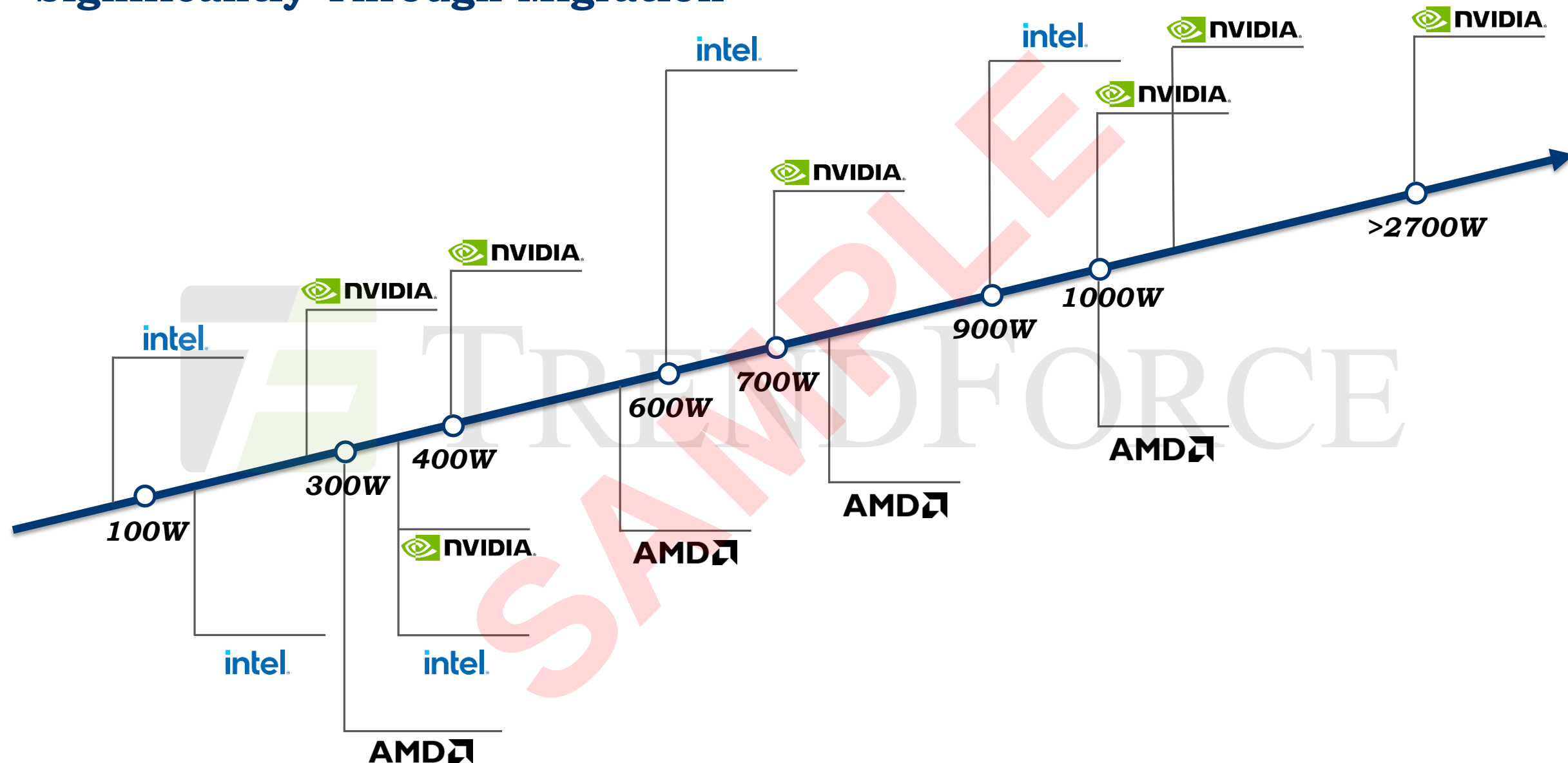
- Google and AWS are the earlier and most proactive CSPs in developing their own ASIC AI chips...

CSPs	AI Chip Name	Node	Major Design Partner	Memory Type	Single Chip Capacity	Chip QTY	Total Capacity	Remark
	TPU v5 Inference							
	TPU v5 Training							
	TPU v6 Inference							
	TPU v6 Training							
	TPU v7 Inference							
	TPU v7 Training							
	Inferentia v2							
	Trainium v2/Ultra							
	Inferentia v2.5							
	Trainium v3							
	Maia v1							
	Maia v2							
	MTIA v1							
	MTIA v2							

Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

AI Chip's Thermal Design Power (TDP) Has Increased Significantly Through Migration



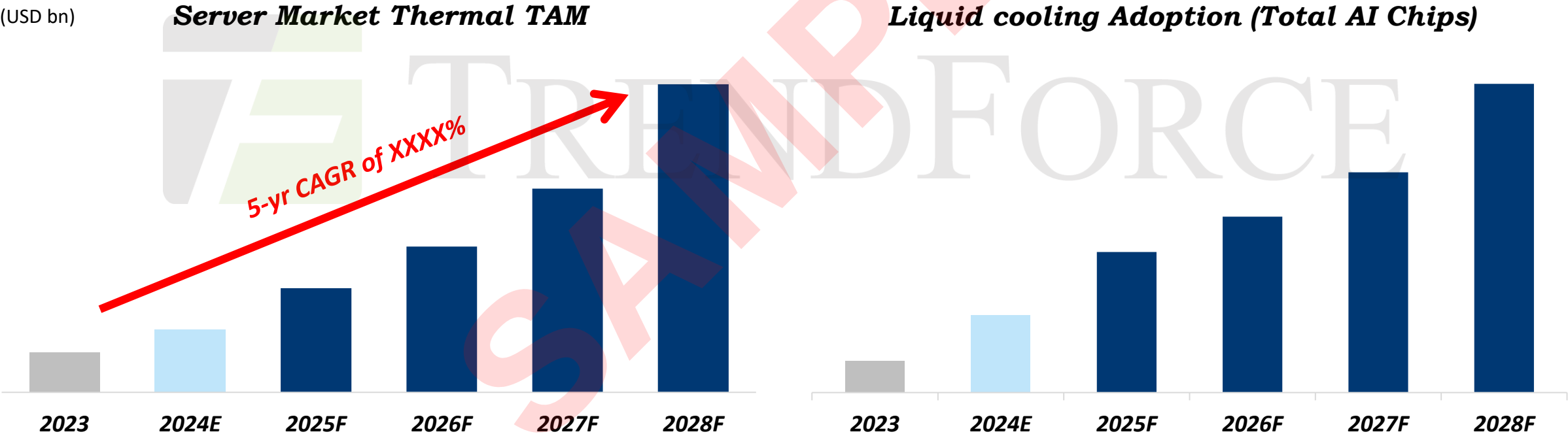
Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

FOR INTERNAL USE ONLY

Liquid-Cooling Adoption Projected to Surge to XXXX% by 2028 alongside Growths of AI Servers

As computing power increases, liquid-cooling solutions are becoming increasingly essential at the data center level for effective heat dissipation and system cooling...

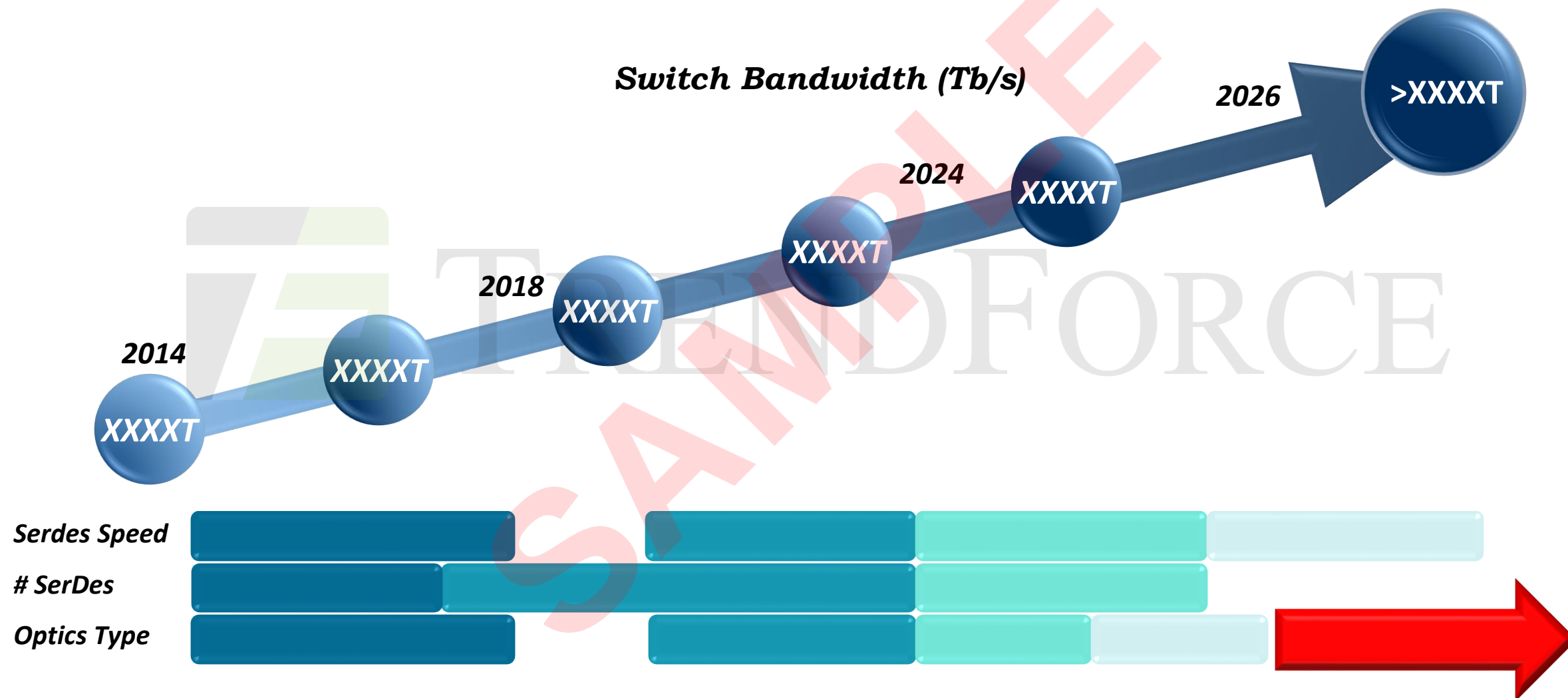


Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

CPO Technology Adoption is Inevitable as Data Transmission Speed Expands

□ The increasing demand for faster speed, higher bandwidth...



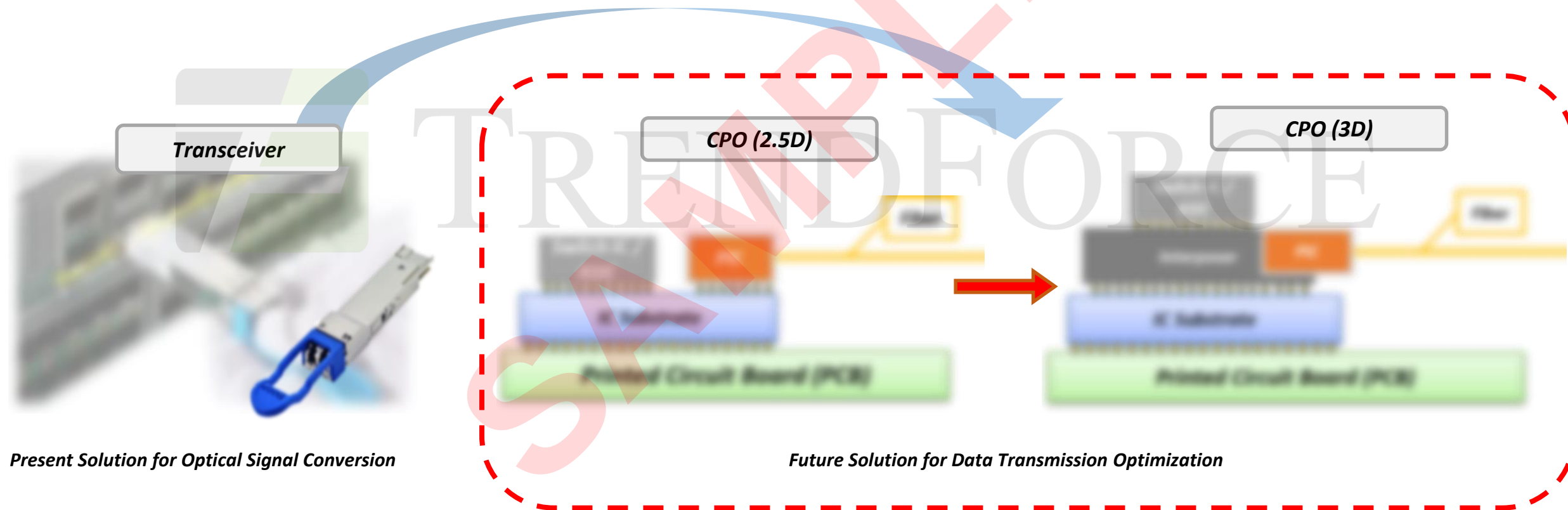
Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

Transformative Shift from Pluggable Optics to CPO to Address Issue of Signal Attenuation

- Signal attenuation issue also intensifies as data transmission accelerates in speed...

Transceiver Encounters Signal Attenuation Challenges



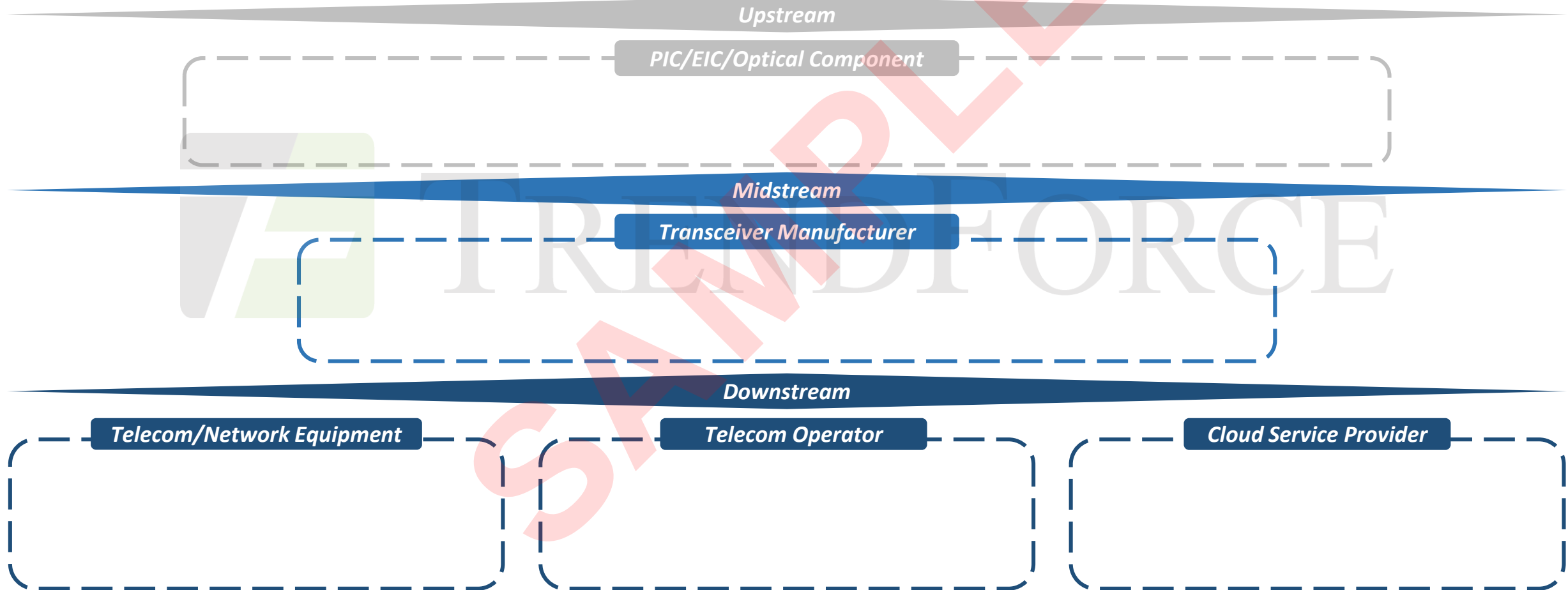
Present Solution for Optical Signal Conversion

Future Solution for Data Transmission Optimization

US Suppliers Maintain a Significant Presence in Optical Transceiver Market

- The optical module supply chain is evolving towards higher performance, diversification, and cost efficiency.

Optical Transceiver Supply Chain



Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

1	XXXX
2	XXXX
3	XXXX
4	XXXX
5	XXXX

Source: TrendForce, Dec. 2024

The contents of this report and any attachments are confidential and legally protected from disclosure.

Thank You

To find out more about TrendForce, scan the QRCode below.

Or visit the URL for further information.



English
www.trendforce.com



Traditional Chinese
www.trendforce.com.tw



Simplified Chinese
www.trendforce.cn

TrendForce & TRI Research Areas



Sales & Services

Semiconductor Research

DRAM, NAND Flash, Foundry

SR_MI

SR_MI@TrendForce.com

Green Energy Research

Solar PV

GER_MI

GER_MI@TrendForce.com

Optoelectronics Research

Micro LED, Mini LED, VCSEL, UV, Video Wall, Lighting

OR_MI

OR_MI@TrendForce.com

Display Research

TFT-LCD, OLED, Smartphone, Tablet, NB, Monitor/AIO, TV

DR_MI

DR_MI@TrendForce.com

ICT Application Research

Communication & Broadband, Consumer Electronics, Innovative Technological Applications, Automotive, Industry 4.0, Gov. & Ent.

TRI_MI

TRI_MI@TrendForce.com